Automated Verification of Mesoscale Forecasts using Spatial Statistics

David W. Jones
Applied Physics Laboratory, University of Washington
1013 N.E. 40th St.
Seattle, WA 98105-6698

phone: (206) 543-3236 fax: (206) 545-4404 email: dwjones@apl.washington.edu

Scott A. Sandgathe
35574 Camp Creek Rd.
Springfield, OR 97478
phone: (541) 988-0289 fax: (541) 988-0289 email: sandgathe@apl.washington.edu

Caren Marzban
Applied Physics Laboratory, University of Washington
1013 N.E. 40th St.
Seattle, WA 98105-6698
phone: (206) 221-4361 fax: (206) 545-4404 email: marzban@stat.washington.edu

Award Number: N00014-01-G-0460/0049
http://cemos.apl.washington.edu
http://coamps.apl.washington.edu/COAMPS-bin/COAMPSOS_homepage.cgi

LONG-TERM GOALS

The verification of high-resolution mesoscale numerical weather predictions presents unique challenges. Traditional verification metrics – root mean square error, etc., which rely on single point verification often give incomplete or misleading assessments of model performance. Small-scale features are often miss-represented (aliasing) or, due to much lower predictability than large-scale features, cause an unwarranted penalty by conventional verification measures due to small spatial or temporal errors. Both the model developer and the operational user need better metrics in order to assess the performance of very high-resolution models. Our long-term goal is to contribute to better high-resolution model development and selection by developing a suite of verification tools to assist both the model developer and the model user.

Towards this goal we are developing three verification techniques: cluster analysis (Marzban and Sandgathe, 2006, 2008a and Marzban, Sandgathe, and Lyons, 2008); variograms (Marzban and Sandgathe, 2008b); and optical flow (Marzban and Sandgathe, 2008c). These techniques are in various stages of development but all have been discussed in previous annual reports and tested against a three model data set of three different WRF formulations: University of Oklahoma's 2km WRF (arw2), NCAR's 4km WRF (arw4), and NOAA's Mesoscale Model (nmm4).

We are now refining these techniques for possible transition to operational usage and preparing to demonstrate them on a year-long data set of the Penn State Mesoscale Model (MM5); the NCAR

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
30 SEP 2008	**Annual**	**00-00-2008 to 00-00-2008**

4. TITLE AND SUBTITLE	
Automated Verification Of Mesoscale Forecasts Using Spatial Statistics	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Washington,Applied Physics Laboratory,1013 N.E. 40th St.,Seattle,WA,98105-6698**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S)
	11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Code 1 only

14. ABSTRACT
The verification of high-resolution mesoscale numerical weather predictions presents unique challenges. Traditional verification metrics ? root mean square error, etc., which rely on single point verification often give incomplete or misleading assessments of model performance. Small-scale features are often miss-represented (aliasing) or, due to much lower predictability than large-scale features, cause an unwarranted penalty by conventional verification measures due to small spatial or temporal errors. Both the model developer and the operational user need better metrics in order to assess the performance of very high-resolution models. Our long-term goal is to contribute to better high-resolution model development and selection by developing a suite of verification tools to assist both the model developer and the model user.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**7**	

Weather Research and Forecast Model (WRF) and the Navy Coupled Ocean and Atmosphere Prediction System (COAMPS) all operating at the University of Washington.

OBJECTIVES

The initial objective of this research effort was to develop the variogram approach as a tool for improving our other mesoscale verification technique (MVT), namely the "box size" (or characteristic length scale as discussed in Hoffman et al 1995 and Nehrkorn et al., 2003). We have modified that objective to explore variograms as a verification tool in their own right. We are still pursuing MVT; however, we have now shown that some components of the MVT are closely related to what is known in image processing circles as Optical Flow (OF). OF appears to lay down a theoretical/mathematical foundation for MVT which adds insight into the inner-workings of MVT. We have pursued the OF approach to verification and have applied it to the above-mentioned forecasts (i.e., arw2, arw4, and nmm4).

WORK COMPLETED

At last report, we had developed all three verification methods at a level of maturity that warranted their application to a large data set consisting of 32 days of obervation and forecasts from three different NWP formulations. In doing so, it became clear that the variogram method must be extended to allow for the computation of variograms for only nonzero grid values. The task was nontrivial, because such grid values constitute an irregular grid. That work is now complete. The value of such a variogram is in its ability to capture only intensity errors. By contrast, the original variogram, computed across the entire grid, assesses intensity and phase errors (size and displacement). In fact, in the application of the method to large data sets, we are computing both type of variograms – one for assessing forecast quality in terms of total error (phase and intensity), and another for intensity error only. All of this has been described in a revised version of an article (Marzban and Sandgathe 2008b), which is now conditionally accepted for publication, pending minor revision.

As for the optical flow method, it has been applied to 4 days consisting of large winter storm events. Only sea-level pressure is examined thus far, because the continuity of the field is more consistent with the assumptions underlying the method. The approach leads to two quantities – 1) a 2d field displaying the vectors that best map the observed to the forecast field (i.e., called the OF field), and 2) a joint histogram of the magnitude and direction of the vectors across an OF field. The former provides a visual display of the quanlity of the forecasts, while the latter offers a summary of the comparison. For example, a peak in the joint histogram would suggest a coherent relative shift of the two fields. The methodology has been extended to allow the averaging of the OF field across multiple days. Details of the approach appear in article which is conditionally accepted for publication, pending minor revision (Marzban and Sandgathe 2008c).

RESULTS

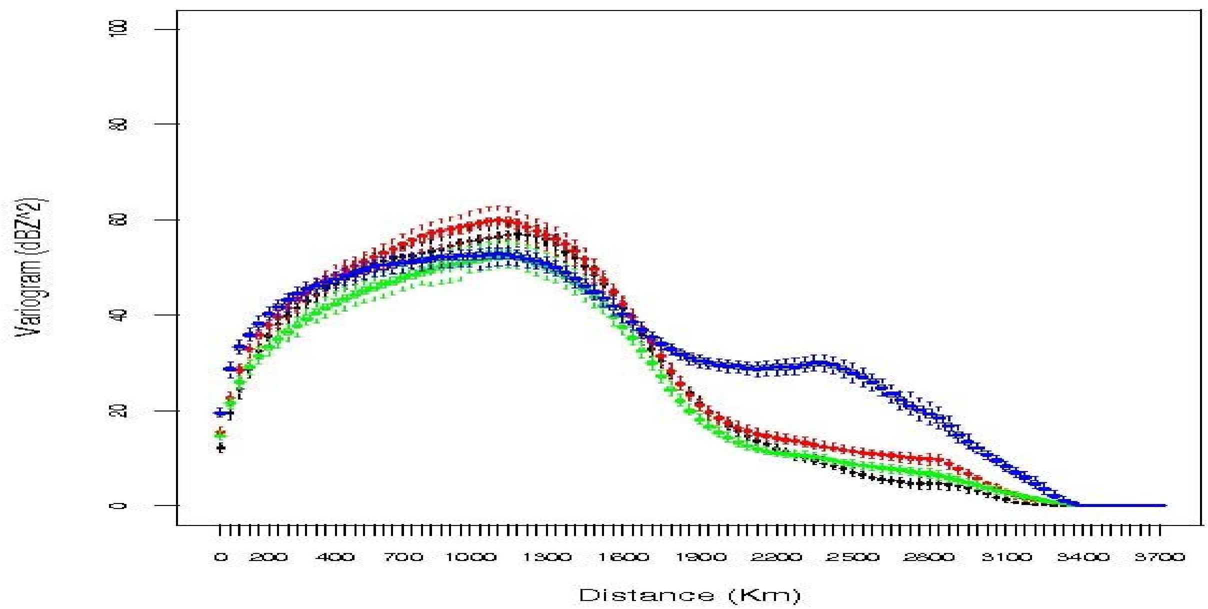
Figure 1 shows both variograms - across entire field (top), and across only nonzero grid values (bottom) - of the observed field (black), arw2 (red), arw4 (green), and nmm4 (blue), all for May 13, 2005. From the left panel it can be seen that on shorter scales (e.g., less than 2000Km), the 3 models have variograms that mostly agree with that of observed field, at least after taking into account the error-bars. On longer scales (exceeding 2000km), although arw2 and arw4 agree with the observed variogram, nmm4 is significantly higher. This suggests that on larger scales, nmm4 does not "decorrelate" as quickly (as a function of scale) as arw2, arw4, and the observations. The right panel assesses only intensity errors, and tells a different story: Most notably, even on smaller scales the variogram for nmm4 under estimates that of the observed field. arw2 and arw4 continue to provide reasonable estimates of the observed variogram; the exception is on larger scales, where none of the models properly represent the observed variogram.

The comparison of the variograms for the observed and forecasts fields (across entire field) for all 32 days in the data, suggests that arw2 and arw4 have reasonable forecasts across all scales, but nmm4 is inconsistent with observed variograms across all scales. However, in terms of intensity error alone (i.e., using variograms computed across only nonzero grid values), it turns out that all three formulations are statistically equivalent with each other and equally consistent with observations. The models produce variograms that are slightly lower than observed, but the difference is not statistically significant.

Figure 2 shows the OF field relating the forecast sea-level pressure to that observed for Dec. 4, 2007. Whereas a visual comparison of the forecast and observed fields (top row) does not immediately highlight areas of disagreement, the OF field (bottom row) clearly shows the differences. Such OF fields and their joint histograms have been computed for 4 days (not shown), and the average OF field across the 4 days is shown in Figure 3. The spatial distribution of the errors is clearly highlighted.

We are participating in an international verification exercise hosted by the National Center for Atmospheric Research (NCAR). They have established a set of idealized cases and also a standardized data set of 5 days observations/forecasts from the 2005 Spring Experiment. We have tested our three techniques on this comprehensive data set for comparison with other mesoscale verification techniques. In addition, we have started using the University of Washington Short Range Ensemble Forecast (SREF) system (MM5 and WRF) in combination with data from the APL-UW COAMPS-OS prediction system. We presented a verification using both variograms and OF of four Northwest winter storm events at the March Pacific Northwest Weather Workshop.

In March, we began archiving UW SREF MM5 and WRF data and APL COAMPS-OS data in order to process a large data set using our three verification techniques. We are currently testing the three techniques on all three model formulations on at least eight months of continuous forecasts.



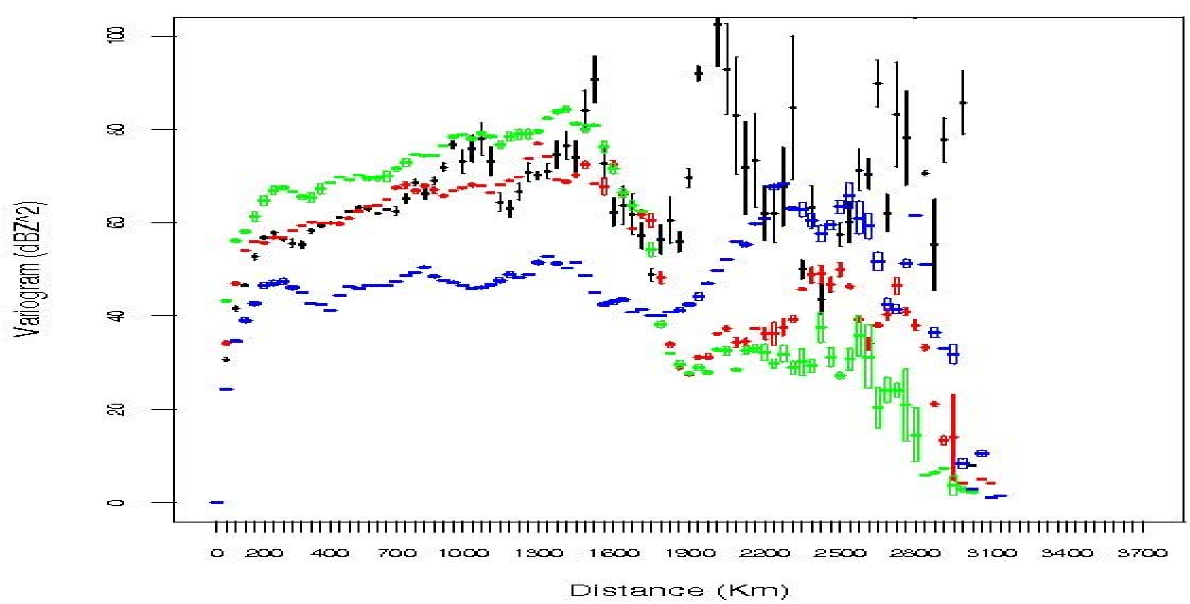


Figure 1: Variogram of the observed reflectivity field on May 13, 2005 (black), and the correspondin g 24hr forecasts from arw2 (red), arw4 (green), and nmm4 (blue). Top variogram is computed across entire field, while bottom variogram excludes zero grid values.

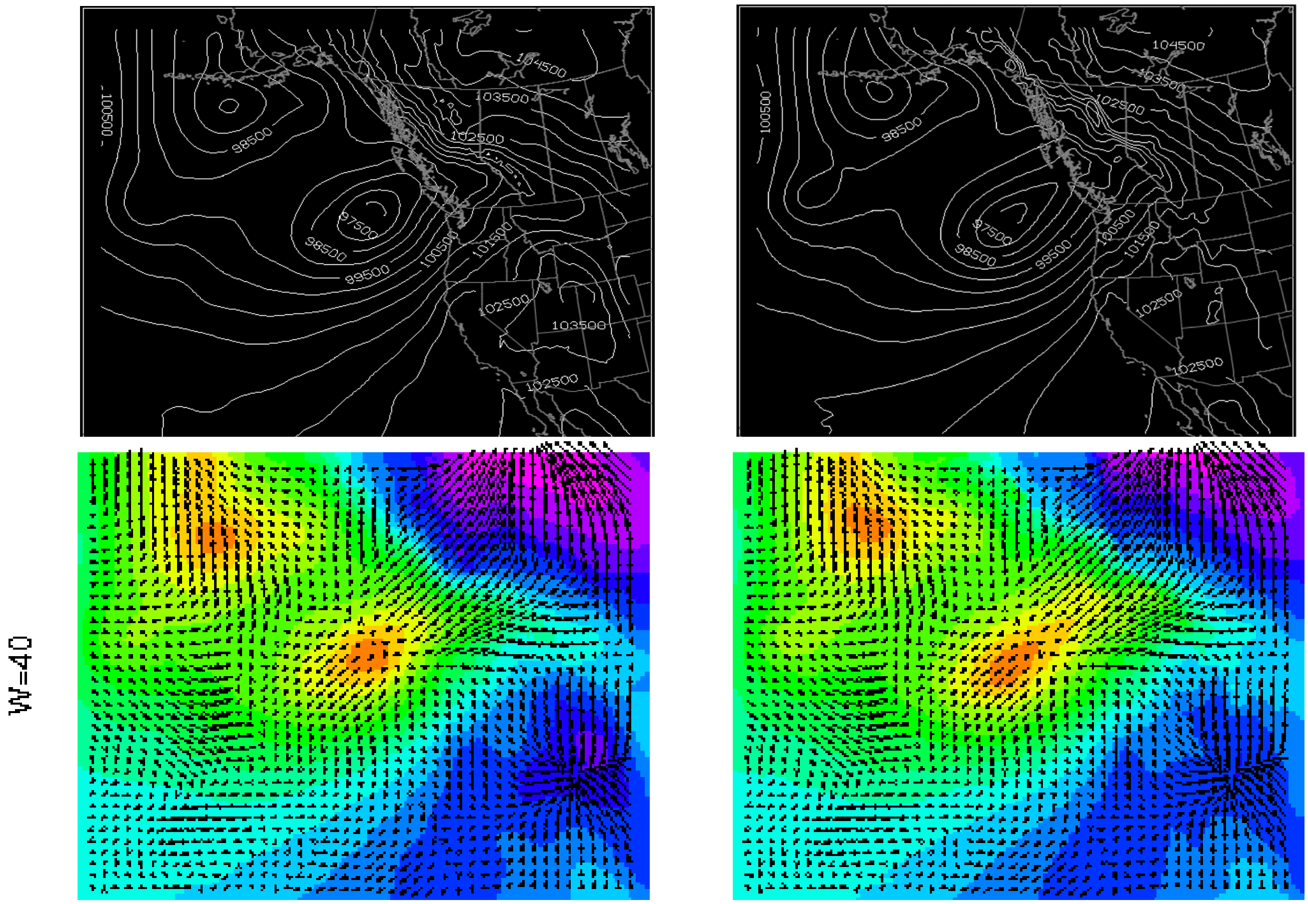


Figure 2: Sea-level pressure forecast (top, left) and observed (top, right) fields for Dec. 3, 2007. Also shown are the optical flow field superimposed on the two fields (bottom row).

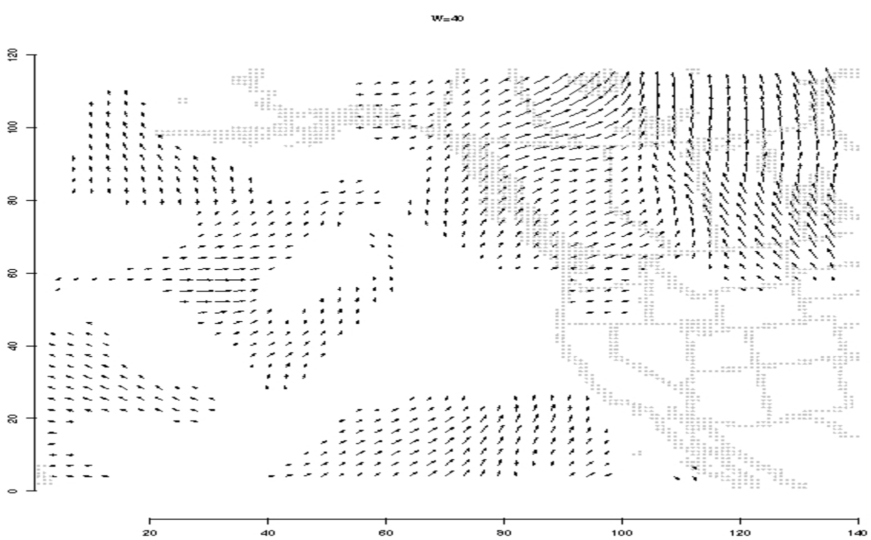


Figure 3: The optical flow field for sea-level pressure averaged over 4 days

CURRENT/FUTURE WORK

In regards to the COAMPS-OS system, Figure 5 is an example of a twelve-hour forecast output field at 15km resolution.

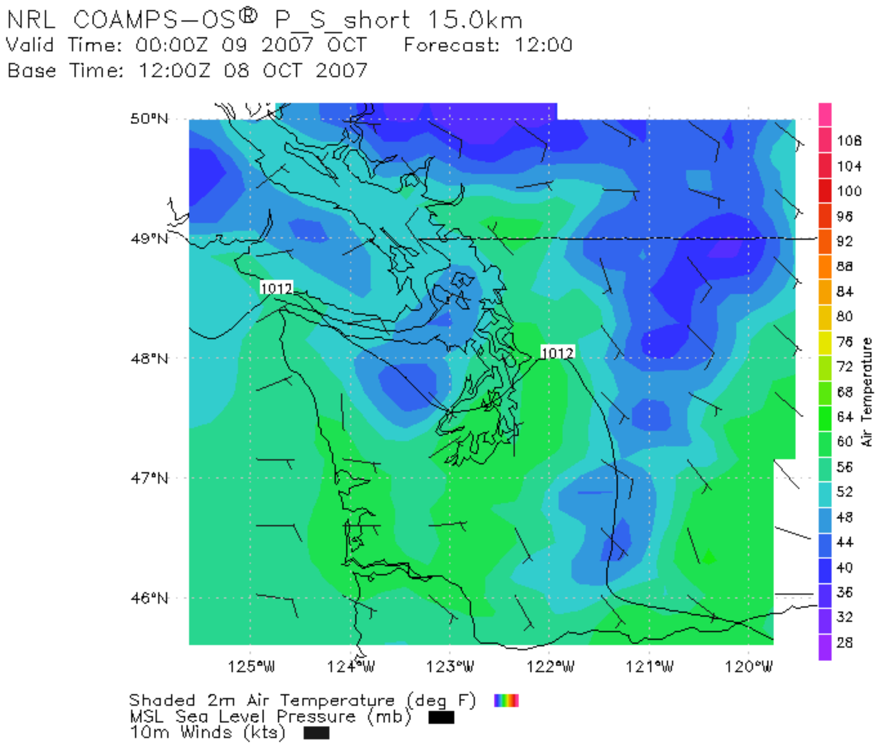


Figure 5. COAMPS-OS 12 hour forecast of air temperature, MSL sea level pressure and 10m winds. The domain is centered over the northwest of the United States.

RELATED PROJECTS

A parallel effort is a verification techniques based on statistical cluster analysis. It is funded by NSF for rapid transition into the DoD/NOAA/NCAR mesoscale WRF Developmental Testbed Center (DTC). The funding completed in August, 2008. It has been applied to precipitation fields (Marzban and Sandgathe, 2006), a variation on that theme has been applied to reflectivity fields (Marzban and Sandgathe, 2008a), and an improved version (that allows the incorporation of meteorological parameters into the analysis) has been reported by Marzban, Sandgathe, Lyons (2008).

PUBLICATIONS

- Reilly, C., P. Price, A. Gelman, S. Sandgathe, 2003: Using image and curve registration for measuring the goodness of fit of spatial and temporal predictions. Journal of Biometrics, 60, 954-964.
- Marzban, C., and S. Sandgathe, 2006: Cluster analysis for verification of precipitation fields. *Wea. Forecasting*, **21**, **5**, 824-838.

- Joslyn, S., K. Pak, D. Jones, J. Pyles, and E. Hunt, 2007: The Effect of Probabilistic Information on Threshold Forecasts. *Wea. Forecasting*, **22, 10**, 804-812.
- Marzban, C., and S. Sandgathe, 2008a: Cluster Analysis for Object-Oriented Verification of Fields: A Variation. Mon. Wea. Rev., 136, 1013-1025.
- Marzban, C., and S. Sandgathe, 2008b: Verification with Variograms. Conditionally accepted at Mon Wea. Rev.
- Marzban, C., and S. Sandgathe, 2008c: Optical Flow for Verification. Conditionally accepted at Wea.& Forecasting.
- Marzban C., S. Sandgathe, and H. Lyons, 2008: An Object-oriented verification of three NWP model formulations via cluster analysis: An objective and a subjective analysis. Mon. Wea Rev., 136, 3392-3407.